

Lesson 029

Hypothesis Tests

Friday, November 17

Decision Making with Data

- Often we want to make decisions using the results of data that have been collected.
- This often comes down to a question regarding the comparison of parameters, either to one another or to some existing standard.
 - Does the new process improve yields over the previous?
 - Can the material withstand the necessary stress conditions?

Hypothesis Testing

- A **hypothesis test** provides one avenue for making these decisions.
- In statistics, a **hypothesis** is a statement regarding the value of a parameter(s).
 - For instance, $\mu = 10$ or $\mu_1 \geq \mu_2$.
- Hypothesis tests use collected data to weigh evidence on competing hypotheses, to infer what is probable.

The Null Hypothesis

- The **null hypothesis** captures our **default belief** about the world.
- We will assume that it is true unless we have evidence to the contrary.
- The null hypothesis needs to be carefully selected, and it is designed to be conservative.
- This will be denoted H_0 and it always has to have an equality sign, either $=$, \geq , or \leq .
 - Often will be something like "no effect", $H_0 : \theta = 0$.

The Alternative Hypothesis

- The **alternative hypothesis** is the claim that we are testing to see if it is supported by the evidence.
- It is always defined in contrast to the null.
- Without evidence, we will act as though our alternative hypothesis is false.
- Typically denoted H_A or H_1 , and does not have an equality.
 - H_0 with $=$ implies H_1 with \neq .
 - $\geq \implies <$ and $\leq \implies >$

If we have $H_0 : \theta \geq 5$, what is the corresponding alternative hypothesis?

$H_1 : \theta < 5$

0%

$H_1 : \theta = 5.$

0%

$H_1 : \theta > 5.$

0%

$H_1 : \theta \geq 5.$

0%

If p_A and p_B are parameters corresponding to the proportion of defects produced by processes A and B , how would we test if the processes behave the same?

$H_0 : p_A \geq p_b$ vs. $H_1 : p_a < p_b$.

0%

$H_0 : p_A \neq p_b$ vs. $H_1 : p_a = p_b$.

0%

$H_0 : p_A = p_b$ vs. $H_1 : p_a \neq p_b$.

0%

$H_0 : p_A \leq p_b$ vs. $H_1 : p_a > p_b$.

0%

Batches of a manufactured chemical are discarded if they contain more than $\delta\%$ of trace contaminants. What null hypothesis should be used to see if a new manufacturing procedure is acceptable?

$$H_0 : \theta = \delta.$$

0%

$$H_0 : \theta \leq \delta.$$

0%

$$H_0 : \theta \geq \delta.$$

0%

$$H_0 : \theta \neq \delta.$$

0%

Important: you must select your hypotheses, both null and alternative, before you've observed your data.

The Level of Significance

- With your null and alternative hypothesis, you need to decide what level of evidence is required to convince yourself of the truth of the alternative.
- Extraordinary claims require extraordinary evidence.
- Consider the importance of a sound conclusion and the stakes if you get it wrong.
- Ultimately, we need to choose a value $\alpha \in (0,1)$.

The Level of Significance (cont.)

- The α will be equivalent, in a sense, to the one selected for confidence intervals.
- Smaller values of α require **stronger** evidence.
- Roughly speaking, α corresponds to how **unlikely** we think that the null needs to be, to act as though it is untrue.
- People typically select $\alpha = 0.05$ and move on. Don't.

Which of the following would require the smallest value of α .

Testing the safety of a new drug to be administered to humans.

0%

Comparing the fuel efficiency of two engine designs.

0%

Ranking the abilities of draft eligible hockey prospects.

0%

Resolving a bet with a friend over the bias of a coin.

0%

Test Statistics

- With a level of significance, null, and alternative hypotheses, we need a test procedure.
- When forming confidence intervals we considered a quantity $T(\theta, \hat{\theta})$ with a known sampling distribution.
 - We refer to this as a **test statistic** and can use it to test our null hypothesis.
- If our null hypothesis is $H_0 : \theta = \theta_0$, then we want to consider how likely it is to observe $T(\theta_0, \hat{\theta})$.

P-Values

- Once we observe the sample, we can compute a value $T(\theta_0, \hat{\theta}) = t$.
- We can ask "what is the probability we would observe something this extreme, if the null hypothesis is true?"
- This is the **p-value** $p = P \left(T(\theta_0, \hat{\theta}) \geq t \right)$
- If the null is true, then this is the probability we observe an extreme result by random chance.

Interpreting P-Values

- If p is very small then the event is **very unlikely** assuming that the null hypothesis were true.
 - As a result, it is better to assume that it is not.
- The p-value is **not**:
 - The probability that the null hypothesis is false.
 - The probability that we make an error in the test.
 - The probability that the alternative is true.

A test statistic is computed as $t = 4.25$, with a corresponding p-value of 0.04. Which of the following are true?

The probability that we observe $t = 4.25$ is 0.04.

0%

The probability that we observe something as extreme as $t = 4.25$ is 0.04.

0%

Assuming the null hypothesis were true, the probability we observe $t = 4.25$ is 0.04.

0%

Assuming the null hypothesis were true, the probability we observe something as extreme as $t = 4.25$ is 0.04.

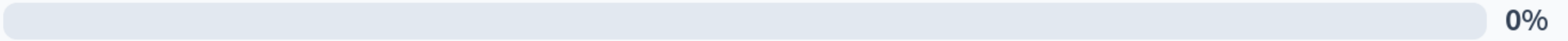
0%

More than one statement is correct.

0%

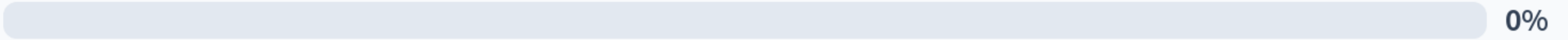
When do you want your Problem Set #2 to be due?

Wednesday November 22nd, @ Midnight



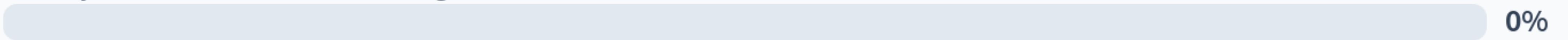
0%

Thursday November 23rd, @ Midnight



0%

Friday November 24th, @ Midnight



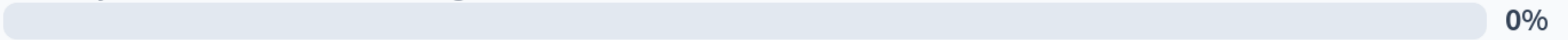
0%

Saturday November 25th, @ Midnight



0%

Sunday November 26th, @ Midnight



0%

Suppose we calculate a p-value as 0.08. What can we say?

The probability that the null hypothesis is false is 0.08.

0%

The probability that the null hypothesis is true is 0.08.

0%

The probability that our conclusion is wrong is 0.08.

0%

The probability that the alternative is true is 0.08.

0%

None of the above

0%

Drawing Conclusions: Fail to Reject H_0

- If $p \geq \alpha$, we did not observe sufficient evidence to reject the null hypothesis.
- We "fail to reject the null" and continue to act as though it were true.
- We never "accept the null" hypothesis (we don't "prove innocence").
- There was not strong enough evidence in the sample.

Drawing Conclusions: Reject H_0

- If $p < \alpha$, we observed sufficient evidence to reject the null hypothesis.
- We "reject the null" in favour of the alternative.
- We act as though the alternative were true.
- Smaller values for α require even stronger evidence in order to reject the default state of affairs.

Hypothesis Testing with Critical Values

- Using critical values we can solve for t_α such that $P(|T(\theta_0, \hat{\theta})| \geq t_\alpha) = \alpha$.
 - Then, if we observe t such that $|t| > t_\alpha$, we must have that $p < \alpha$.
 - If we observe t such that $|t| \leq t_\alpha$, we must have that $p \geq \alpha$.
 - For normal distribution this will be $t_\alpha = |Z_{\alpha/2}|$.
- This allows for hypothesis testing without p-values.

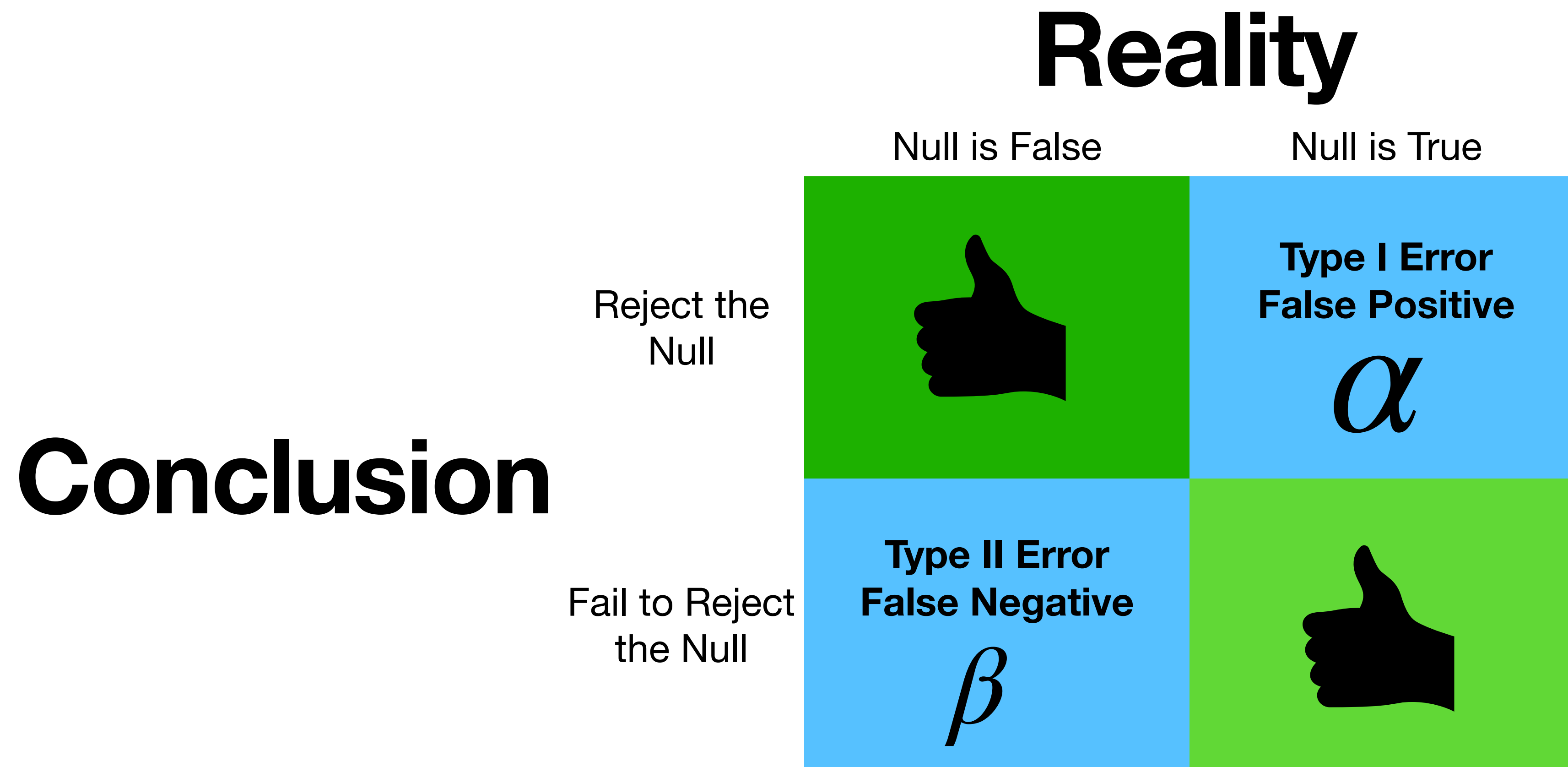
Hypothesis Testing: Procedure

1. State the null and alternative hypotheses.
2. Decide on your level of significance, α .
3. Compute the test statistic.
4. Find the p-value (or critical value).
5. Compare p to α (or t to t_{α}).
6. Draw conclusions.

Drawing Conclusions

- If we "reject the null hypothesis" we say that the result is **statistically significant** (at level α).
- If we "fail to reject the null hypothesis" we say that the result is **not statistically significant** (at level α).
- A reported p-value indicates how strong the evidence is against the null hypothesis.
- We can be wrong.

Errors in Hypothesis Testing



Errors in Hypothesis Testing

- The probability of a type I error is α , the level of significance.
- The probability of a type II error is β .
 - We call $1 - \beta$ the power of the hypothesis test.
- We want α and β to each be as small as possible, but they trade-off against one another.
 - Typically we set α then try to minimize β .

Suppose we are testing $H_0 : \theta = 3$, and find a p-value of 0.02. Which of the following is true?

We reject the null hypothesis at a 0.05 level of significance.

0%

We fail to reject the null hypothesis at a 0.01 level of significance.

0%

We have that $|t| \geq t_{0.03}$.

0%

We have that $|t| = t_{0.02}$.

0%

All of the above

0%

Suppose that when testing $H_0 : \theta \geq 3$ we find a p-value of 0.02 and so we reject the null hypothesis. Which of the following is true?

We know that $\theta < 3$.

0%

We may have made a type I error.

0%

We may have made a type II error.

0%

None of the above

0%